Misclassification of the response variable

Chriel, M.*
Danish Cattle Association, Frederiks Allé 22, DK-8000 Aarhus C, Denmark.
Tel: +45 33 91 44 46
Fax: +45 33 32 35 03
E-Mail: mac@vetinst.dk

Abstract
Advanced statistical and epidemiological methods are widely used in clinical studies. This initiative can only be welcomed. However, the risk of introducing misclassification in observational studies carried out in collaboration between clinicians and modellers is high. Misclassification can be introduced by a number of sources.

Going through a number of published papers on equine colic has demonstrated that even an obvious simple classification of fatalities and survivors can cause trouble. The consequence of the misclassification will lead to lack of reliability of the statistical models in general, along with invalid and between studies variable results and conclusions.

Introduction

Advanced statistical and epidemiological methods are widely used in clinical studies. The correct diagnosis and treatment of acute abdominal pain in equines is important for the survival. Information on the pre-operative clinical and clinicopathological findings is essential in the decision between medically treatable cases and cases which require surgical intervention (Parry et al, 1983,). However, due to the urgent nature for treatment of the condition and the often very expensive horses, owners request a precise answer regarding the diagnosis and the probability for survival and returning to normal exercise. Therefore, several models to predict the diagnosis and prognosis have been developed (Thoefner et al, 2000, Furr et al, 1995, Reeves et al, 1990, Pascoe et al, 1990, Ducharme et al, 1989) by using advanced statistical methods – focusing on the choice of a statistical method for a decision support system.

Potential sources of misclassification

Generally, bias is known to originate from selection of study objects, incorrect information collected on the study subjects and/or failure to adjust for the variables other than the study factor that are predictive for the outcome (Kleinbaum et al, 1982). Non-differential error of a binary disease outcome will usually produce bias towards the null, representing no exposure-disease association, provided that the misclassification is independent of other errors. Non-diffentiality of measurement errors is far from sufficient to guarantee that bias will be toward the null. Such guarantees require that the exposure errors also be independent of errors in the other variables, including disease and confounders, a condition that is not always plausible. Often, despite imperfect information on the study object, it is often possible to assess the direction of the distortion but not the magnitude of the bias. But, if the

misclassification is severe enough, the bias can completely obliterate an association and even reverse the direction of association.

The classical Berkson´s fallacy (Berkson, 1946) is well known where the studied population differs from the general population making inferences invalid. Having access to only a small number of observations, involvement of large number of clinicians and the acute condition often complicates clinical studies of equine colic. The sensitivity and specificity of the diagnostic tests, and the multicollinearity between the clinicopathological and diagnostic parameters complicate the analyses. Methods for adjusting for multicollinearity and combining the risk factors have been developed – like factor analysis, principal component analysis, and correlation analysis (Dohoo et al, 1996) but a substantial number of observations are necessary.

An example from equine colic studies

Studies carried out on equine colic have revealed that even an apparent simple classification of fatalities and survivors can cause misclassification (Ducharme et al, 1983, Orsini et al, 1988, Reeves, et al, 1989, Reeves,et al 1990; Furr et al, 1995). Problems can occur in the classification when the underlying diagnosis is not accounted for. When the prognostic value of various clinicopathological variables is evaluated for the relevance in a decision support system for horses having surgical treatment (e.g. torsion) the case must be classified as a fatality if the exploratory celiotomy diagnosed a strangulating lesion despite survival and discharge from the hospital after successful surgical correction of the underlying condition. Other studies have corrected part of the misclassification by re-classifying the fatalities and euthanized horses by taking the underlying condition into consideration after post mortem examination but not corrected the classification of the survivors (Thoefner et al, 2000).

Subsequent advanced multivariable statistical analyses with formation of descriptive causative factors for the survival of the admitted horse can cause difficulties in interpretation of the components (Thoefner et al, 2001). Furthermore, misclassification of the response variable can be part of the explanation of the very different clinical and clinicopathological parameters reported in various studies (see Reeves et al, 1989) – and presumably in combination with substantial variation between veterinarians in the sensitivity of the applied clinical diagnostic tests.

Concluding remarks

There is a increasing use and understanding of the importance of applying advanced statistical methods for analysis and interpretation of data from clinical studies. The quality of the results is however completely relying on the quality of the data. Misclassification cannot absolutely be avoided. A thorough discussion of the definition of the response variable must take place with relevant statistical modellers in order to ensure application of correct methods. The consequence of the misclassification will lead to lack of reliability of the statistical methods in general, along with invalid and misleading results and conclusions.

References

Adams, S.B., McIlwraith, C.W., 1978: Abdominal Crisis in the Horse: A Comparison of Pre-Surgical Evaluation With Surgical Findings and Results. Vet Surg 7:63-69

Berkson, J., 1946: Limitations of the application of fourfold table analysis to hospital data. Biomet Bull 2:47-53.

Dohoo, I.R., Ducrot, C., Fourichon, C., Donald, A., Hurnik,D., 1996: An overview of techniques for dealing with large numbers of independent variables in epidemiologic studies. Prev Vet Med 29:221-239

Ducharme, N.G., Hackett, R.P., Ducharme, G.R., Long, S., 1983: Surgical Treatment of Colic Results in 181 Horses. Vet Surg 12: 206-209

Ducharme, N.G., Pascoe, P.J., Lumsden, J.H., Ducharme, G.R., 1989: A computer-derived protocol to aid in selecting medical versus surgical treatment of horses with abdominal pain. Equine Vet J 21:447-450

Furr, M.O., Lessard, P., White II ,N.A., 1995: Development of a Colic Score for Predicting the Outcome of Equine Colic. Vet Surg 24:97-101

Kleinbaum, D.G., Kupper, L.L., Morgenstern, H., 1982: Epidemiologic Research Principles and Quantitative Methods, New York, ISBN 0-534-97950-5

Orsini, A., Elser, A.H., Galligan, D.T., Donawick, W.J., Kronfeld, D.S., 1988: Prognostic index for acute abdominal crisis (colic) in horses. Am J Vet Res. 49:1969-1971

Parry, B.W., Gay, C.C., Andersson, G.A., 1983: Assessment of the necessity for surgical intervention in cases of equine colic: A retrospective study. Equine Vet J 15:216-221

Pascoe, P.J., Ducharme, N.G., Ducharme, G.R., Lumsden, J.H., 1990: A Computer-derived protocol Using recursive Partitioning to Aid in Estimating Prognosis of Horses with Abdominal Pain in Referral Hospitals. Can J Vet Res 54:373-378

Reeves, M.J., Curtis, C.R., Salman, M., Hilbert, B.J. 1989: Prognosis in Equine Colic Patients Using Multivariable Analysis. Can. J Vet Res 53:87-94

Reeves, M.J., Curtis, C.R., Salman, M., Reif, J.S., Stashak, T.S., 1990: A Multivariable prognostic model for equine colic patients. Prev Vet Med 9:241-257

Reeves, M.J., Curtis, C.R., Salman, M., Stashak, T.S., Reif, J.S., 1991: Multivariable prediction model for the need for surgery in horses with colic. Am J Vet Res, 52:1903-1907

Thoefner, M.B., Ersboell, A.K., Jensen, A.L., Hesselholt, M., 2001: Factor analysis of the interrelationships between clinical variables in horses with colic. Prev Vet Med 48:201-214

Thoefner, M.B., Ersboell, A.K., Hesselholt, M., 2000: Prognostic indicators in a Danish hospital-based population of colic horses. Equine Vet J, Suppl., 32:11-18